DOCUMENT RESUME

ED 478 326 SE 068 157

AUTHOR	Liebovitch, Larry S.; Todorov, Angelo T.; Wood, Mark A.; Ellenbogen, Kenneth A.
TITLE	When Using the Mean is Meaningless: Examples from Probability Theory and Cardiology.
PUB DATE	2000-00-00
NOTE	15p.; In: Kelly, Anthony E. and Richard Lesh, Handbook of Research Design in Mathematics and Science Education, Mahwah, NJ, Lawrence Erlbaum, 1999. p913-26.
AVAILABLE FROM	Lawrence Erlbaum Associates, Inc., 10 Industrial Avenue, Mahwah, NJ 07430. Web site: http://www.erlbaum.com.
PUB TYPE	Books (010) -- Reports - Research (143)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	*Curriculum Development; *Educational Research; Elementary Secondary Education; *Mathematics Education; Research Design; *Research Methodology; *Statistics

ABSTRACT

This chapter describes how the mean of fractal processes does not exist and is not a meaningful measure of some data. It discusses how important it is to stay open to the possibility that sometimes analytic techniques fail to satisfy some assumptions on which the mean is based. (KHR)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 326

32 When Using the Mean Is Meaningless: Examples From Probability Theory and Cardiology

Larry S. Liebovitch
Angelo T. Todorov
Florida Atlantic University

Mark A. Wood
Kenneth A. Ellenbogen
Medical College of Virginia

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. Barker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abraham Maslow (1966), writing on the psychology of science, noted that, "If the only tool you have is a hammer, you tend to treat everything as if it were a nail." The corollary of this proposition is that sometimes we drive nails, and sometimes we smash windows. The statistical tools that we use to analyze data each carry with them certain assumptions about the characteristics of the data. We use these methods so often that we can easily forget these assumptions. Frequently, we have concentrated so much on teaching how to use these methods that we may not have made clear the assumptions on which these methods are based, or we may not even know those assumptions. When the data satisfy these assumptions, then the analysis that we perform can provide faithful models of the phenomena that we wish to describe. Otherwise, the analysis may be worthless, or dangerous, and worse yet, we may not even question our methods. Like a chess player, we must look not only at the piece our opponent just moved, but also remain aware of the move's strategic implications. We can easily forget that a method as simple, as commonplace, as seemingly innocuous, as taking the arithmetic mean of a set of numbers depends on certain assumptions. In this chapter, we see

SE068157

BEST COPY AVAILABLE

how important it is to stay open to the possibility that our analytic technique fails to satisfy those assumptions.

ASSUMPTIONS ON WHICH THE MEAN IS BASED

We measure something in nature. We repeat the measurements a number of different times. By combining these separate measurements, we hope that the errors in each measurement will cancel and the result will be more accurate. A way to do this is to compute the arithmetic mean. We have computed so many means in our life, and seen the means computed by so many other people, that this seems the natural, and perhaps even the only, way to handle the data. It seems almost absurd to question the value of the mean. Yet, the usefulness of the mean depends on the data satisfying certain assumptions. Surprisingly, many experimental data do not meet these assumptions! The means computed from such data can lead to a false picture of what the data represent.

The assumption that underlies the use of the mean is that a set of numbers have about the same value, with some a bit smaller and others a bit larger, as shown in FIG. 32.1a. In this case, the mean provides a good estimate of the best guess of a typical value. The probability density function (PDF) characterizes the number of times each value is found in the data. More technically, the probability that we find a value x, in the range $(x, x + dx)$ is given by PDF(x). Usually, we assume that the PDF is Gaussian, or asymptotically Gaussian. That is, it has the form commonly called a bell curve or a normal distribution. The mean of such a Gaussian PDF is well defined and easy to compute. The best guess (maximum likelihood estimator) of the mean of the PDF is the arithmetic mean of the data.

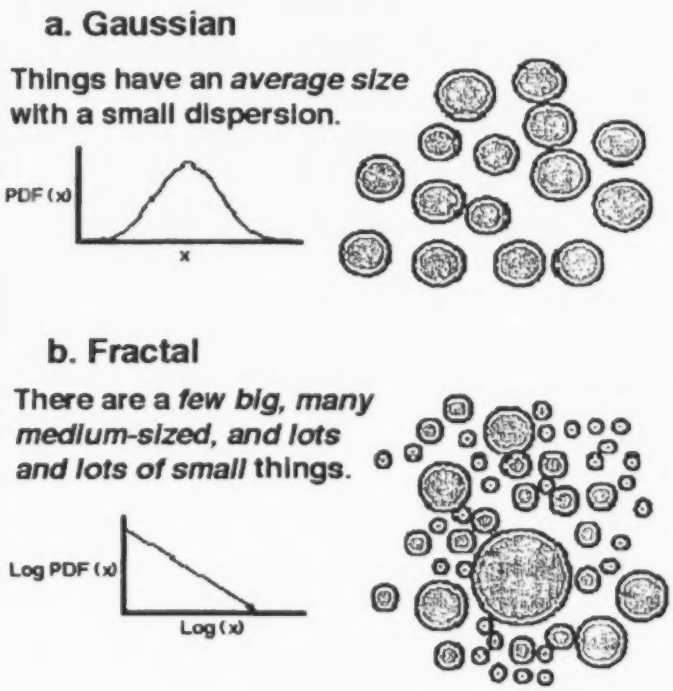
a. Gaussian
Things have an average size with a small dispersion.
PDF (x)
x
b. Fractal
There are a few big, many medium-sized, and lots and lots of small things.
Log PDF (x)
Log (x)

FIG. 32.1. (a) We are used to thinking that a set of similar objects, such as these circles, have approximately the same diameter, with some a bit smaller and others a bit larger. The mean of the diameters of these circles is a good way to characterize the size of these objects. The probability that a circle has a diameter in the range x, $x + dt$ is given by the probability density function, PDF(x). The PDF is a Gaussian distribution, which is characterized by its mean and variance. (b) Many things in the natural world are fractal. They consist of an ever larger number of ever smaller objects. The mean is not defined and is not a good way to characterize the size of these objects. The PDF is a power law and thus a straight line on a log-log plot.

The mean value of data sampled from the population is called the sample mean. As we include ever more data, the fluctuations in the values around the typical value cancel out, and the sample means approach a finite limiting value. We consider this limiting value the "true" value of the population, which we define as the population mean. Technically, the central limit theorem guarantees that as we collect more data, the values of the sample means

approach that of the population mean. Our faith in the central limit theorem may lead us to believe that it applies equally to all phenomena. We may assume that sample means of all data converge to a finite limiting value.

THE MEAN IS NOT DEFINED FOR FRACTALS

Data drawn from many processes and phenomena in the natural world do not satisfy the assumptions underpinning the central limit theorem. Many objects in nature can be described as consisting of a larger number of ever finer pieces, as shown in FIG. 32.1b. Typically, there are a few big pieces, many smaller pieces, and a really large number of tiny pieces. Objects with these properties are called "fractals" (Bassingthwaighte, Liebovitch, & West, 1994; Mandelbrot, 1983). For example, the blood vessels in the body consist of a few large arteries, many smaller arterioles, and a huge number of tiny capillaries. This characteristic description can be applied to the branching patterns of trees, the distribution of particles in the soil, the clustering of galaxies, the timing between failures in computer components, and many other objects in space or processes in time in the natural world (Mandelbrot, 1983).

For these fractals, as we include more data or examine more pieces at finer resolution, the sample means do not approach a finite limiting value. For example, for some fractals, as we include more data, the increasing number of smaller pieces dominates the mean, and consequently the mean approaches zero. For other fractals, as we include more data, the small number of very large pieces dominates the mean, and consequently the mean increases without bound; that is, it approaches infinity. Because the sample means do not approach a finite limiting value, the population mean is not defined!

What happens if we do not realize this, and we try to use the mean of a set of measurements to characterize a fractal? The value of the mean depends on the amount of data analyzed. If we measure the same fractal object twice, but use different amounts of data, we will find two different values of the mean. If we measure two similar fractal objects, but use different amounts of data, we will find two different values of the mean. Thus, there are important consequences to our failure to realize that the usual interpretation of the mean depends on certain assumptions that are not satisfied by the data from fractals. These consequences can affect life-and-death decisions, as we see later in the analysis of the time

intervals between episodes of abnormal heart rhythms. These intervals have fractal properties and as a result, cannot be characterized by a mean value. Therefore the state of one patient, or a group of patients, and the consequences of their medical therapies cannot be assessed by the changes in the mean value of these time intervals.

HISTORICAL BACKGROUND

Sometimes, pure mathematics is motivated by the need to solve worldly problems. About 400 years ago, mathematicians began to study games in order to determine the odds that gambling houses should offer their customers. This was the beginning of probability theory and, subsequently, statistics. About 250 years ago, Niklaus Bernoulli, in Russia, created a game with a PDF with fractal properties, which was published by his cousin Daniel Bernoulli and became known as the St. Petersburg Paradox (Dutka, 1988).

In the 1920s, the French mathematician Paul Lévy analyzed these distributions, which became known as stable or Lévy stable distributions (Mandelbrot, 1983). In these PDFs, the parameters of the sum of independently distributed random variables are linear combinations of the parameters of the individual distributions (Feller, 1971). The standard Gaussian distribution is a subset of these stable PDFs. However, these stable PDFs also include more general distributions.

One way to characterize any PDF is by its moments. The moments are the values of variables, raised to a power, and weighted by their frequency of occurrence in the PDF. For example, the first moment is the mean, the average of all of the values in the data set. The second moment is the variance, which characterizes the spread in the values around that average value. The mean and the variance are the most common methods used to analyze and characterize a set of any measurements. The mean and the variance are well defined for the standard bell curve, that is, the Gaussian PDF. However, the moments, such as mean and variance, do not necessarily exist for stable PDFs. The properties of these stable PDFs are different from those predicted by the central limit theorem, which assumes that the second moment exists and is finite.

Mandelbrot developed the properties of fractals further and introduced them to the wider scientific community (Mandelbrot, 1983). Although the properties of the stable PDFs of

fractals were well known to a small group of mathematicians, these properties were not known to most people in the natural sciences. That is why these properties appear so strange at first sight. We have been too used to dealing only with Gaussian, or asymptotically Gaussian, distributions. With the realization that many objects and processes in nature have fractal properties, understanding the properties of stable PDFs and how they are different from Gaussian distributions is essential in order to analyze and understand data from the many fractals that are found in nature. We now illustrate these ideas with two examples of fractals: one from probability theory and another from the world of medicine.

AN EXAMPLE WHERE THE MEAN EXISTS: A GAUSSIAN COIN GAME

Consider the game that consists of one toss of a coin. If the coin lands heads up, we win \$1; if the coin lands tails up, we win \$0. The expected winnings is the probability of each outcome multiplied by the winnings of that outcome. The probability that the coin lands heads up is 1/2 and the probability that the coin lands tails up is 1/2. Thus, the expected winnings of one game is equal to (1/2)(\$1) + (1/2)(\$0) = \$0.50. In a fair gambling house, the house should charge the player \$0.50 to play each game, and the player would accept this as a fair charge.

If we play the game many times, in the long run, the winnings, averaged over all of the games, should be the same as the expected winnings of one game, which is \$0.50. The more games that are played, the closer the mean winnings of all of the games will be to \$0.50. That is, as more games are played, the sample means computed from all of those games approach \$0.50. This game can be simulated with a computer. FIG. 32.2a shows the average winnings after playing N games as a function of N. The average winnings after N games is the sample mean for N games. As expected, as more games are played, the average winnings after N games approaches \$0.50 rapidly.

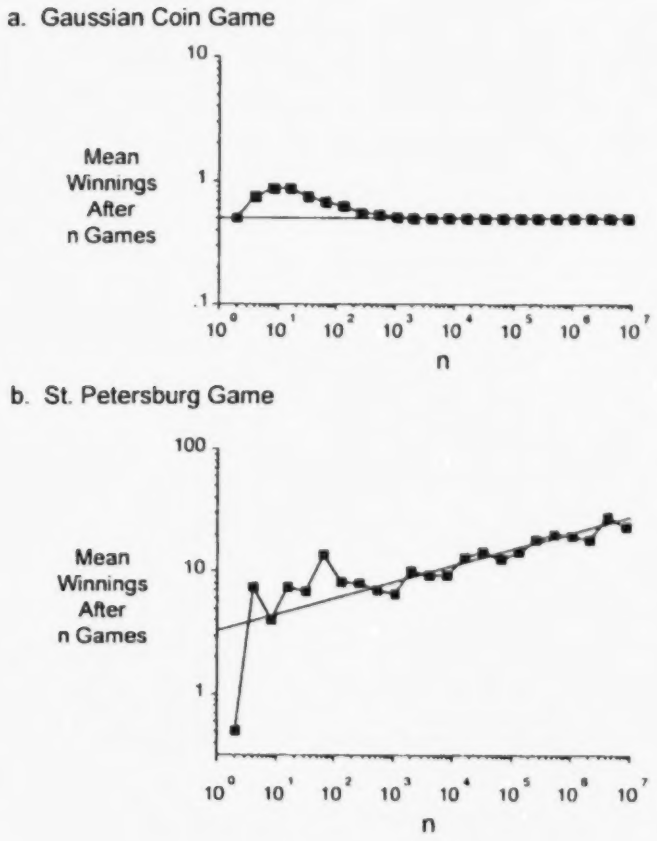

FIG. 32.2. (a) In the Gaussian coin game, as the number of games N played increases, the mean winnings after n games approaches a finite limit (\$.50) that we identify as the population mean. (b) In the fractal St. Petersburg coin game, as the number of games N played increases, the mean winnings after N games increases without bound. Therefore, the population mean is not defined.

We can compute the PDF of the winnings for this game. Let us say that we play a set of 100 games. The PDF(x) is the probability that x, the mean winnings, that is, the winnings of all of those games divided by the number of games, is between x and $x + dx$. The most probable value of the mean winnings x is \$0.50. It is less probable that the mean winnings will be a little lower or a little higher than \$0.50. It is very less probable that the mean

winnings will be much lower or much higher than \$0.50. Thus, the PDF has a peak at the mean winnings of \$0.50 and decreases as the mean winnings increases or decreases from \$0.50. Consequently, the curve of the PDF has the characteristic shape of a bell. Therefore, it is called a bell curve, which is known also as a normal, or a Gaussian, distribution. FIG. 32.3a shows the PDF, this Gaussian curve, computed for a set of 100 such games.

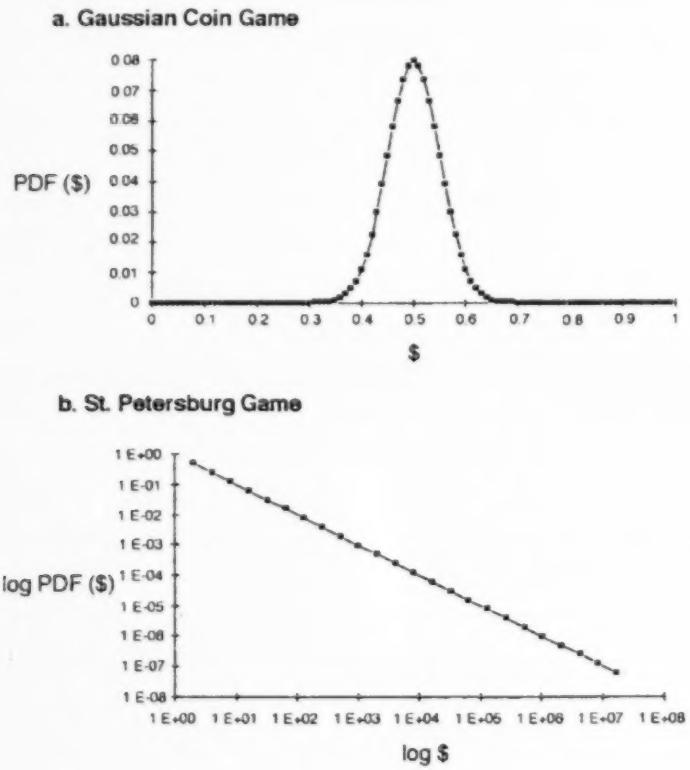

FIG. 32.3. (a) The PDF of the Gaussian coin game is a Gaussian distribution. Shown here is the distribution computed for 100 games. (b) The PDF of the fractal St. Petersburg game is a power law distribution and thus a straight line on a log-log plot. (Note computer printout notation 1 E + a = 1.0 x 10**a.)

AN EXAMPLE WHERE THE MEAN IS NOT DEFINED: THE ST. PETERSBUG PARADOX, A FRACTAL COIN GAME

Now let us play the game formulated by the Bernoullis. In this game, we toss a coin as many times as needed until we get a head. If the coin lands heads up on the first toss, we win \$2. If the first toss is a tail, and then on the second toss we get a head, we win \$4. If the first and second tosses are tails, and then on the third toss we get a head, we win \$8. If the first, second, and third tosses are tails, and then we get a head on the fourth toss, we win \$16, and so on. That is, if we get a head on the m-th toss, we win $\$2^m$. The probability of getting a head on the first toss is 1/2. The probability of getting a tail on the first toss and then a head on the second toss is 1/4. The probability of getting a tail on the first two tosses and then a head on the third toss is 1/8. The probability of getting a tail on the first three tosses and then a head on the fourth toss is 1/16, and so on. That is, the probability of getting a head on the m-th toss is 2^{-m}. The expected winnings of one game are (1/2)(\$2) + (1/4)(\$4) + (1/8)(\$8) + (1/16)(\$16) + . . . = infinity! This game became known as the St. Petersburg Paradox because the house wants to charge the player more than all the money in the world to play even one game, whereas the player feels he or she should be charged only \$2 to play because he or she has a 50-50 chance of winning that amount on one play of the game.

We can simulate this game on a computer. FIG. 32.2b shows the average winnings after playing N games as a function of N. The average winnings after N games is the sample mean for N games. The more games are played, the larger the mean winnings! The mean does not approach a finite limiting value. There is no limiting value that we can identify as the population mean. The value that we find for a sample mean will depend on the number of games played.

We will find different values for the mean winnings evaluated after different numbers of games. Our experience with Gaussian statistics suggests that when the moments of a data set are constant in time, then the rules that generated the data must also be constant in time: when the moments vary in time, then the rules that generated the data must also vary in time. For example, the most commonly used moment is the mean. If the mean is constant in time, we think that the process that generated the data must be constant in time. When the mean varies in time, we think that the process that generated the data must be changing in time. With only

our Gaussian thinking cap on, we would think that these changing means imply that the rules of the game are changing over time. But the rules of the St. Petersburg game are not changing over time. Our experience with Gaussian statistics does not give us the right intuition for dealing with stable PDFs.

The probability PDF(2^n) of winning \$$2^n$ is equal to 2^{-n}. The PDF as a function of the amount won is a straight line when log[PDF(2^n)] is plotted versus log(2^n). This is shown in FIG. 32.3b. This power law form is characteristic of the stable PDFs of fractals. It means that as this game is played again and again, very often the winnings are small, sometimes the winnings are larger, and very rarely the winnings are huge. This pattern of geometric repetition of an ever smaller number of ever larger winnings is characteristic of a fractal.

The infinite value of the mean in this game (the expected winnings) troubled those who were developing probability theory (Dutka, 1988). Yet, this game is as reasonable and as logical as the first coin toss game described earlier with a finite mean. Somehow, though, the infinite value of the mean of this game and its paradox did not seem like a "fair" game. Probability problems with these characteristics were still studied by some mathematicians over the next 250 years, but PDFs with these characteristics were not considered in the main thread of probability theory that became popular in the natural sciences. That is why the properties of this game seem so odd. But, they are not odd at all. The many fractal objects in nature have these fractal properties. We must be able to recognize and interpret these properties properly when we find them in nature, if we are to understand the natural world, as shown in the following example.

AN EXAMPLE FROM HEART DATA

Sudden cardiac death is the single, most common cause of death in the United States. It almost always arises from abnormalities in the rhythm of the beating of the heart. The heart normally contracts in an organized wave that pushes blood out of the ventricles, the main pumping chambers of the heart. When each small segment of the heart beats separately, the wall of the heart twitches like a bag of worms and the blood is not pumped. This condition is called ventricular fibrillation and it leads to death within minutes. Work is underway to determine what other abnormal rhythms of the heart presage the onset of ventricular fibrillation. This

work may make it possible to predict the onset of fibrillation and perhaps prevent it by the timely use of drugs or electrical stimulation. One dangerous rhythm that can lead to ventricular fibrillation is called ventricular tachycardia, or "v-tach" in the medical jargon. In v-tach, the ventricles beat over three times their normal rate. It is now possible to implant small computers in the chest, called cardioverter defibrillators, that monitor the heart rate. When an episode of v-tach is detected, these devices send a strong shock throughout the heart that resets it to a normal rhythm. These computers can record when each episode occurs. These times can be read back later by radio communication with a device that triggers the computer to replay the information that it has stored.

If one did not know anything about fractals or stable PDFs, one might think that a good way, perhaps the only way, to characterize how a heart patient is doing is to determine the mean time between episodes of v-tach. One might also think that measuring how the mean time changed when different drugs were taken would be a good way to determine the effects of the drugs. This is exactly what the medical community had thought.

We measured the PDF of the times between events of v-tach in 30 patients, recorded by their implanted cardioverter defibrillators (Liebovitch et al., 1999). For 28 patients, these events were infrequent enough that we had to combine the data from different patients in order to have enough event times to determine the PDF. For two patients, there were enough events of v-tach to determine the PDF from each patient separately.

As shown in FIG. 32.4, the PDFs from all of these heart patients have the same form. That is, the plot of the log(number of events in the range t to $t + dt$) versus log(t) is a straight line. This power law form is characteristic of the stable PDF of a fractal. The computation of the PDF and its presentation on a log-log plot have revealed a new and striking form that would not have been revealed by the computation of the sample mean, or even by a linear plot of the PDF.

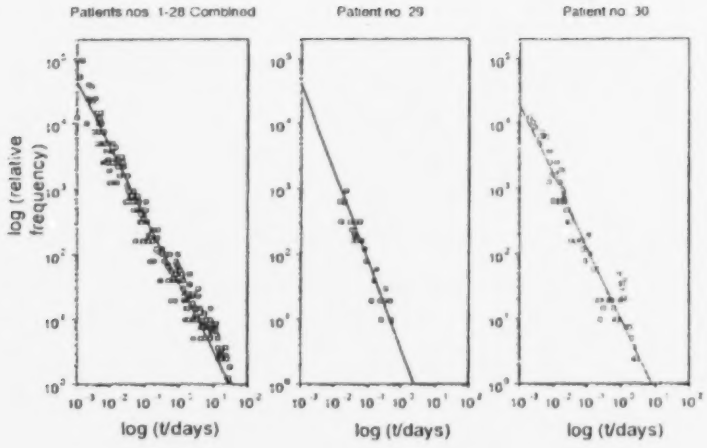

FIG. 32.4. The relative frequency, which is proportional to the PDF, of the times t between episodes of abnormally rapid heart rhythms (ventricular tachycardia). These PDFs have the power law form of the stable PDFs of fractals, like that 32.3b. The mean time between episodes is not defined. The time between episodes is often very brief, sometimes longer, and rarely very long. The mean fails to properly characterize these data. However, the true nature of these data is revealed clearly by the straight lines on these log-log plots of the PDFs.

The simple, clean form of the power law of the PDF in the data combined from 28 different patients is most likely due to the fact that each patient also has a PDF of the same power law form. If the PDF from each patient were different, then combining the data from different patients would produce an unrecognizable form of the PDF. This conjecture is supported by the fact that the two additional patients (Nos. 29 and 30) each have the same power law PDF as that of the combined data.

This power law PDF extended over an impressive range of five orders of magnitude in times between v-tach events from seconds to days. This range of five magnitudes is one of the largest ever reported for a power law PDF in biomedical data. This observation is remarkable considering that these patients have different types of heart disease and different medical treatments. These data may be revealing a completely unanticipated property of the how the heart works.

The central lesson is that the mean is not an appropriate method to characterize these data. There is no mean value that can characterize the time between episodes of v-tach. These times can be very long, more often they are short, and most often they are very brief. The form of the PDF, revealed as a straight line on a log-log plot, reveals, for the first time, the true nature of the timing of these life-threatening events.

This striking finding would not have been revealed if we had determined only the mean time between events of v-tach. In fact, because there are different amounts of data from each patient, the mean is different for each patient. This also shows that determining the mean time between events of v-tach, before and after a drug is given, will not be a meaningful way to test the effectiveness of drugs. Here then is a concrete example where the mean fails because the data have fractal, rather than Gaussian, characteristics. Not understanding the properties of these data could lead us to a meaningless, and maybe even a life-threatening, conclusion.

Another irregularity of the heart beat occurs when the larger ventricles beat before the smaller atria. These beats are called premature ventricular contractions, or PVCs in the jargon. It is not known when PVCs become dangerous. Many people without severe heart disease were given drugs to eliminate their PVCs. Unfortunately, these drugs may have caused more harm than good (Moore, 1995). A decade ago, it was found that the variation in the frequency of these PVCs increased as the larger amount of data over a longer time period was analyzed (G. Schmidt et al., 1988). This behavior surprised the researchers who discovered it. Illuminated by the light of our knowledge of stable PDFs, this suggests that the second moment, the variance, is increasing without bound as it is measured over longer time intervals. This indicates that even healthy people can be expected to have PVCs when observed over long time periods. Perhaps here, too, log-log plots of the PDF may make it possible to characterize the normal rates of occurrence of PVCs and thus determine when they are harmless and should be left alone, and when they are dangerous and should be treated.

SUMMARY

Patterns in our data may be clarified or obscured by the mathematical tools we choose. Each mathematical method makes certain assumptions about the data. We must remain aware of those assumptions and whether our data satisfy them. Even commonplace tools, such as the

arithmetic mean, make assumptions about the data. When the data do not satisfy those assumptions, our analysis may not be meaningful. The worst case is when we do not know that the results are meaningless.

In this chapter, we have described how the mean of fractal processes, such as the winnings in the St. Petersburg game or the timing between episodes of dangerously rapid heartbeats, does not exist and is not a meaningful measure of such data.

When the mean is not meaningful, how then should we analyze such data? For the many fractals found in nature, the meaningful measure is how the mean depends on the amount of data analyzed or the resolution used to analyze it. Technically, there is a scaling relationship where the logarithm of the mean is a linear function of the logarithm of the resolution used to measure it. This shows up dramatically as a straight line on a log-log plot of the PDF. The slope of this line is related to a parameter called the fractal dimension. We are exploring whether the fractal dimension can be used to characterize the degree of risk of life-threatening heart rhythm problems and assess the ability of different medical therapies to lower that risk.

So, mathematics from hundreds of years ago has brought new insight into a current problem, enabling us to see things in a different way, which may lead to saving lives.

ACKNOWLEDGMENT

This work was supported in part by Grant EY6234 from the National Institutes of Health.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").